# XCS: The Data Analysis Pipeline and X-Ray Redshifts

K. Sabirli[1], A.K. Romer[2], M. Davidson[3], C.A. Collins[4], M.J. Hilton[4], S.T. Kay[2], K.R. Land[5], A.R. Liddle[2], R.G. Mann[3], R.C. Nichol[1], M.J. West[6], P.T.P. Viana[7]


## Abstract

The XMM Cluster Survey (XCS; Romer et al. 2001) is currently being assembled on the serendipitous detection of galaxy clusters in X-ray images taken by the XMM-Newton satellite. It aims to derive robust cosmological parameters from measurements of the evolution of the cluster number density and to provide an unprecedented description of cluster scaling relations (see related talk by Viana). The XCS already covers more area than the various ROSAT PSPC serendipitous cluster surveys to a flux limit of at least 1E-14 erg/s/cm$^2$ (0.5-2.0 keV). In this poster we present our fully automated data analysis pipeline. The pipeline begins with the retrieval of public data sets and ends with science quality image products. The pipeline has been designed to meet the various challenges posed by the XMM experiment, e.g. energy dependent vignetting and a flaring proton background. An additional, major challenge involves the need for redshift follow-up. We are, therefore, exploring the applicability of 'X-ray redshifts' to the XCS. Using artificial spectra, we demonstrate the limits of this approach in the presence of a complex particle background.


## Data Analysis Pipeline

An automated pipeline is prepared to retrieve raw public data sets from XMM-Newton Science Archive (XSA), and produce science products to use in the catalogue preparation. It makes use of public tasks such as XMM's Science Analysis Software (SAS), FTOOLS and specifically prepared IDL routines, combined with shell programming. A brief process flow is as follows:

❖ The list of public observations up-to-date is taken from XSA, suitable for the catalogue purposes (i.e. pointings with at least one EPIC camera active during the exposure, with regular camera modes and known filters etc.)

❖ This list is compared with the presently reduced sets, and the list of problematic sets produced using previous XCS logs and XMM pointing logs.

❖ The new pointings that pass the above mentioned criteria get retrieved from XSA, via the java applet and ftp, placed on the production line with initial preparation such as uncompressing.

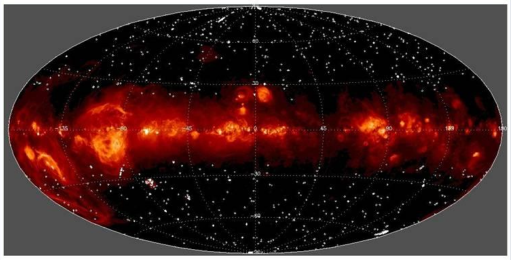
❖ After this step, each pointing is reduced independent of the others. We are working on the merging of the analysis for sets with overlapping regions. At each step, log files are generated to watch for any warning messages. For all pointings added to the processing list the following actions are taken:

✓ Prepare an ODF summary file using *odfingest*.

✓ Check the calibration status and prepare a new calibration summary file using *cifbuild (for consistency, the same calibration sets are used for all sets)* .

✓ For the EPIC camera exposures taken in the imaging mode, linearize the event files using *emproc* for the EMOS cameras, and *epproc* for the EPN camera (not in default mode, some flag values are modified to guarantee the presence of both in and out of field of view (fov, ofov) parts of the cameras, and bad pixel corrections). Patterns [0:12] and [0:4] is selected for EMOS and EPN cameras respectively.

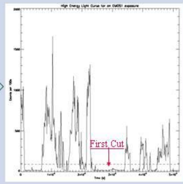
✓ Due to the orbit of XMM, the exposures are generally contaminated with high energy proton flares with varying extends. To clean these, high energy (above 10keV) light curves are generated from event lists with 100 s binning, using fov regions… The rest of the filtering is presented using one EMOS1 exposure:



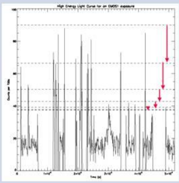
**XCS on the sky in galactic coordinates (~140 square degrees)**

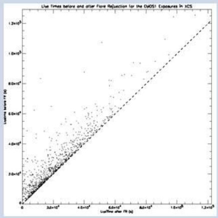
**First, an initial cut :**

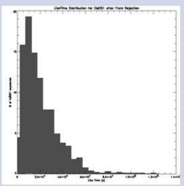
**(1<(counts)<90 for EMOSs and 2 <(counts)<180 for EPN) to clean severe flares.**




**Then, the light curve is cleaned successively using 3σ up cuts until the change in the mean of the counts is negligible. The good time intervals are created for the remaining regions using *tabgtigen to filter the event lists.***







✓ *Evselect* is used to select the good time intervals, and generate images in fov for two energy bands: (0.5-2.0) and (2.0-10.0) keV with ~4" binning.

✓ The images are cut down to 512x512 size and respective exposure maps are generated using *eexpmap*, including the vignetting and quantum efficiency information.

✓ All good exposures for the given pointing are merged and graphic images are produced for visual check.

✓ The sets that pass each check point get their place in the XCS archive and forwarded for multi-resolution source detection (see the poster by Michael Davidson et al.).

✓ After each calibration update, all the pointings are being reprocessed.

## X-Ray Redshifts: Motivation

New X-ray surveys such as XCS and DUO will produce thousands of cluster candidates, but redshifts are required to do cosmological tests (See Liddle, Viana, Romer & Mann, 2001, MNRAS, 325, 875L). Traditional redshift follow-up (optical spectroscopy from the ground) will be almost impossible for these surveys. Follow-up for ROSAT surveys such as e.g. REFLEX & SHARC (<1,000 clusters total) took up to a decade to complete. Automated optical follow-up (e.g. using correlations with the SDSS) is not yet possible over most of the sky. Of the 140 square degrees covered so far by XCS, less than 5 lie in regions covered by public SDSS data. X-ray spectroscopy holds the promise of yielding cluster redshifts directly. Nevalainen et al, 2001, A&A 374, 66 discovered a serendipitous cluster in an XMM pointing and was able to assign a temperature and redshift from the same dataset.

## Methodology 1: Fake Spectra

We created hundreds of thousand artificial XMM cluster spectra using fakeit with XSPEC and the on-axis response matrices. Redshift, temperature, abundance, absorption sampled at random. For each parameter set, spectra with net cluster counts of 500; 1000; 2,000; and 10,000 [0.5-2.0 keV] were generated. Realistic background contamination was added at the 1%, 10%, 100% and 200% level using the Lumb et al. template spectra. Spectra were dumped as ascii files in both folded and unfolded (through telescope response) format and both with and without background subtraction applied.

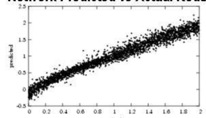
## Methodology 2: Neural Network

We developed a Neural Network with 1 hidden layer. The network was trained using the Casper algorithm on 70% of the available synthetic spectra. 15% of the remaining spectra were used for training. The final 15% were used for validation. The advantage of a network (over XSPEC fitting) is its speed and the fact that it does not require an a priori knowledge of the telescope response function.

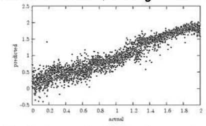
## Results

- Results with low (1% or 10%) background contamination:
  - The network was able to reproduce both the temperature and the redshift very well.
  - Even with only 500 cluster counts (top figure).
- Results with high (100% or 200%) contamination:
  - Almost all the predictive power disappeared.
  - Only the 10,000 count spectra produced acceptable results (bottom figure).

**Network Predicted *vs* Actual Redshifts**




**500 cluster counts, 5 background counts**




**1E5 cluster counts, 2E5 background counts**

## Future Directions

The results from the background contaminated spectra were disappointing. We suspect strong fluorescent lines in the XMM particle background that are causing the neural network to fail. We plan to re-run the tests avoiding those parts of the spectrum. We also plan to run tests applicable to future satellites (those in lower orbits that would suffer less from certain background effects). There remains the possibility that X-ray redshifts will not be practical; future cluster surveys may be severely handicapped by the need for ground based optical follow-up.


**References**

Romer et al., 2001ApJ...547..594R
Liddle et al., 2001MNRAS.325..875L
Lumb et al., 2002A&A...389...93L
Nevalainen et al., 2001A&A...374...66N



**Affiliations**

1. Carnegie Mellon University
2. University of Sussex
3. University of Edinburgh
4. Liverpool John Moore University
5. Imperial College London
6. University of Hawaii at Hilo
7. University of Porto